

Feasibility Study of a Parallel Audio and Video ECC for Watermarks in Short Audio and Video Frames

Kazuhiro Kondo*, Yoshimi Yamaguchi, Toshio Koga

Graduate School of Science and Engineering, Yamagata University
4-3-16 Jonan, Yonezawa, Yamagata 9928510, Japan

*kkondo@yz.yamagata-u.ac.jp

Abstract

We investigated the feasibility of a parallel watermarking method for audio and video signals for multimedia streams with editing and/or switching capability. Since editing and switching are commonly done in short intervals, these watermarks need to be applied on a short frame basis. In this study, the watermark information with error correction coding (ECC) will be split and embedded into audio and video simultaneously. Information decoded from both audio and video signals is to be used collectively to recover the embedded information even if there is severe degradation in the decoded signal. We studied a scheme in which the watermark is embedded in the video stream with ECC, and the ECC is also redundantly embedded into the audio stream. We conducted a very basic simulation using a simple DCT-based watermark for video, when the Bose-Chaudhuri-Hocquenghem (BCH) code is placed in both video and audio streams. At an overall bit error rate of 5%, with BCH (63, 51) ECC, we estimated that the error correction capability can be improved by approximately 35% compared to when all information is embedded into only the video stream. A simple calculation-based estimation assuming uniform error bit distribution also showed error correction capability improvement of approximately 48%, which is slightly higher than it was shown with the simulation.

Keywords

Watermark; Video; Audio; Multimedia; Error Correction Coding

Introduction

Recently, NHK, a leading TV broadcasting network in Japan, jointly with Mitsubishi Electric Corporation, has developed a watermarking system for newsgathering, which is used in their program production, where news coming from many local stations is frequently switched back and forth in short frames (NHK, 2014). Here, each local station is assigned an ID to show which station is sending the newsfeed. The ID, date/time, and other auxiliary

information, if necessary, are represented by approximately 100 bits, and are embedded in the video part as watermarks while still maintaining the original video quality. This information will be used for housekeeping and archival purposes.

Usually, it is relatively easier to embed much more than 100 bits in a short segment of video signals without perceptual quality degradation. On the contrary, it is generally not as easy to embed similar amount in audio signals. However, it is possible to embed 1 bit in approximately 1,000 audio samples, resulting in about 40 bits/s capacity when applied to CD music sampled at 44.1 kHz (Li, 2000), and about 80 bits/s if both of the stereo channels are used. This means that 100 bits of information required for station ID and others, can be actually embedded into either audio or video signals of a few seconds, which is about the minimum time interval for the selection or switching of video sources from one station to another.

Watermarking is widely used in practice. It usually is applied to audio or video signals individually, not in combination. In (Tanaka, 2005), however, the same information is embedded into both video and audio simultaneously. The embedded information is detected by first extracting data from both the video and the audio streams, and then the correlation between these two data streams is calculated. The purpose of this redundancy seems to be to make the watermark extremely robust. However, it is apparent that the capacity of video and audio streams for transparent watermarking is significantly different, and its impact on the perceived quality tends to be greater in audio than video.

In this paper, we investigate the feasibility of a watermark for video and audio, where only the

supplementary error correction code (ECC) is embedded into the audio, while the watermark itself is embedded into the video stream (Yamaguchi, 2012). Since the ECCs usually requires a much smaller number of bits than the information data itself, much less information needs to be embedded into the audio stream compared to the video stream, thereby decreasing the impact on perceived quality. Accordingly, we first define the allocation of embedded information to both the video and audio stream, and estimate the proportion of the frames with bit errors that can be corrected for errors when additional error correction information (ECC) is available from the audio stream.

This paper is organized as follows. In the next section, the parallel watermark method is defined. The following section describes the conditions for the simulation of bit error correction performance by the proposed method, and then the simulation results are given. Then, a description of simple calculations to estimate the ECC performance and its comparison to the simulation results are given. Finally, the conclusion as well as the future plan are given.

Video and Audio Parallel Watermarking for Short Frames

In this section, we define the watermark scheme for video and audio streams. As stated in the introduction, video streams can accommodate significantly more watermark data than its audio counterpart, especially when the quality is kept as transparent as possible. Thus, the watermark itself is embedded into the video sequence. Some ECC is added to this watermark for robustness. We consider this ECC to also be embedded into the audio at the same locations within the stream to enhance the robustness of the watermark.

We consider the following two scenarios for embedding this information into the video and audio streams:

- 1) Embed only the watermark bits in the video; embed ECC in audio only.
- 2) Embed watermark and ECC in the video; embed the same ECC redundantly in audio.

As an extension to 2), we can also consider embedding watermark and ECC in video, and embed the same ECC multiple times redundantly in audio since the ECC is relatively short. We can then combine all ECC and apply majority decision logic, for instance, to

further reduce the probability of errors in the ECC. In this paper, we consider only 2) since the purpose of this study is to grasp the feasibility of this kind of redundancy. The watermark embedding scheme to be studied here is depicted in FIG. 1, where the video and audio is split into the same short time intervals which we call cycles. The watermark and its ECC is embedded into the video in this cycle while the ECC is also embedded in the audio corresponding to the same cycle.

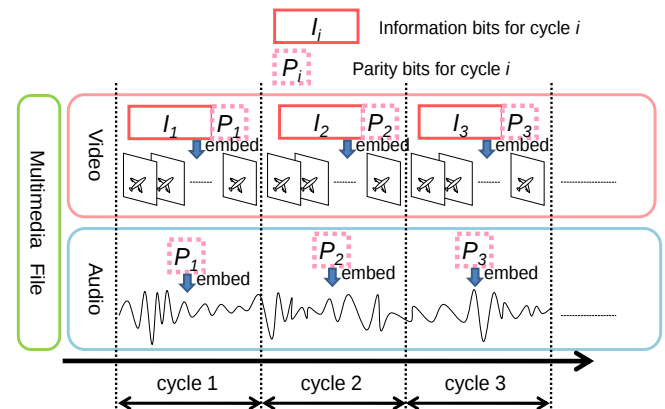


FIG. 1 PROPOSED PARALLEL WATERMARK

Estimation of Robustness Improvement

We now try to estimate the amount of error correction improvement which can be expected by embedding ECC simultaneously in the audio stream compared to embedding both the watermark and ECC in video only.

Error Correction Coding

In this study, we use the simple Bose-Chaudhuri-Hocquenghem (BCH) ECC (Morelos-Zaragoza, 2006) for the purpose of estimation. As stated before, we are targeting applications where the multimedia content is switched in short intervals (cycles), so BCH ECC, which can control the precise number of bits which is corrected and allows correction in short blocks, matches our target well. Most other more powerful ECCs require much longer blocks, and so they may not be the best choices here. We especially use the BCH (63, 51) code which uses 12 parity (ECC) bits for a block of 51 information bits, for a total of 63 bit blocks. This ECC can correct up to two random bits in a block.

Video Watermark

Since the purpose of this paper is to estimate the potential improvement in the robustness of the watermark when data is embedded in both the video and audio streams, a simple and commonly used

video watermarking method is tested. The video input signal is converted to DCT in 8x8 pixel blocks, and the data are embedded in the diagonal components of the DCT block (Piva, 97). FIG. 2 depicts the embedded DCT coefficients. Two predefined independent random bit sequences, one is embed data "0" and the other a "1," are added to the selected components. At the decoder, the watermark can simply be detected by convolving the embedded DCT coefficients with one of the predefined random sequences, the one giving larger correlation values being the detected watermark bit.

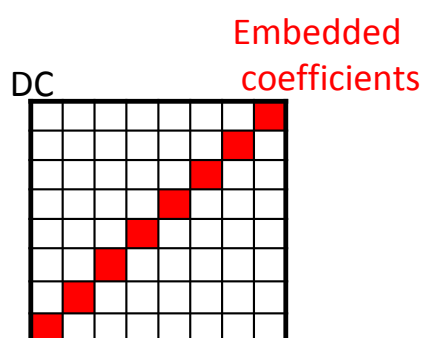


FIG. 2 WATERMARK EMBEDDED VIDEO DCT COMPONENTS

Audio Watermark

A simple watermark using the Direct Spread Spectrum is employed (Boney, 1996; He, 2008). With this method, the watermark bits are spread across all of the bandwidth by multiplying with a predefined m-sequence random bit sequence. The resultant sequence is added to the original audio (host) signal. To decode the watermark, the watermarked signal is correlated with the same m-sequence, and thresholded.

Multimedia Stream

We have selected three multimedia clips from the Consumer Digital Video Library (Consumer Digital Video Library, 2014), shown in FIG. 3, 4 and 5. Video clip 1 is named the "NTIA boy crawling in a ball pit" sequence, clip 2 the "NTIA two geese waking," and clip 3 the "NTIA exterior view of store 2" sequence in the library. All clips are 12 seconds long at a rate of 30 frames per second, 640x480 pixels. All clips have stereo audio, 16 bits per sample, 48k samples per second.

The tested clips vary widely in terms of content and movement. Clip 1 includes many small objects with some movement. Clip 2 includes high frequency contents, but very little movement. Most regions of clip 3 is very dark and almost has no movement.



FIG. 3 CLIP 1 (NTIA BOY CRAWLING IN A BALL PIT)



FIG. 4 CLIP 2 (NTIA TWO GEESE WAKING)



FIG. 5 CLIP 3 (NTIA EXTERIOR VIEW OF STORE 2)

Estimation of Robustness Improvement through Attack Simulation on Video

We embed random bit patterns into the three multimedia clips with the video watermarking described in the previous chapter. The Gaussian filtering attack, a very common form of an attack and easy to apply, is employed here to attack the watermarked clips (Kutter, 1999; Peticcolas, 1999). The Gaussian filter, given by the following equation, is

actually a geometric smoothing filter that attenuates high spatial frequency components in both the horizontal and vertical direction,

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

where $G(x, y)$ is the filter weights for a pixel with horizontal distance x and vertical distance y from the origin (pixel being filtered), and σ is the standard deviation which controls the strength of the filter. This filter is applied as an attack, and attenuates the spatial high frequency included the image. Thus, it degrades the watermark included in the high frequency regions.

In this study, the deviation is set to 3.3, which gives an average watermark bit error rate of 5%. We use this level of attack as the benchmark since the quality of the filtered video is relatively transparent, while with a larger deviation, the degradation becomes quite visible. Within a block of 63 bits (watermark and the ECC), the average number of error bits expected is approximately 3 bits.

We also attempt to watermark audio with the direct spread spectrum method, and attack this with additive Gaussian noise. However, since this method is a very robust method, and the bit rate of the audio watermark is extremely low because we only need to embed the ECC bits, we observe no bit errors unless the attack is so severe that the noise make the audio useless. Thus, we assume that the watermarked audio ECC bits include no bit errors, and the ECC bits are always available from the audio stream intact.

We now calculate the number of blocks that can be corrected with the help of the ECC bits from the audio stream, and which are not possible to correct with the video stream only. As we have stated, for the benchmark condition, 3 error bits per 63 bit block are expected. FIG. 6 depicts the distribution of number of errors for clip 1 with the benchmark Gaussian filter attack.

As we see, the majority of blocks includes 1 to 3 error bits per block (68% in the case of the example in Fig. 6). Single and two bit errors can be corrected with the bits in the video stream only since the BCH (63, 51) ECC can correct up to two random error bits. Three bit errors can also be corrected if at most two error bits are in the information bits, *i.e.*, at least one of the error bits is in the ECC. In this case, the ECC bits in the audio stream, which we can assume to include no errors as stated in the previous paragraph, can be used

to replace the ECC bits in the video stream.

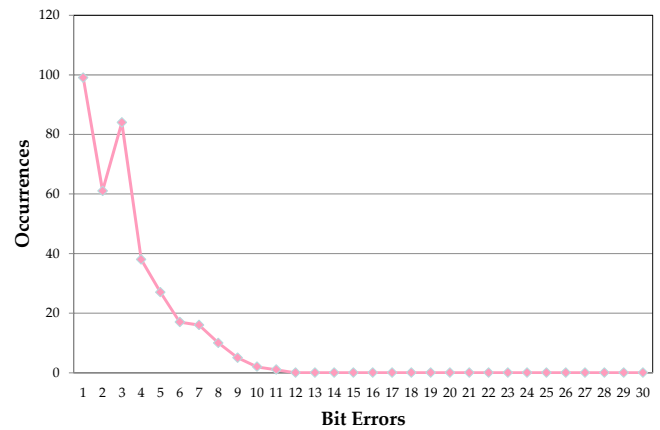


FIG. 6 HISTOGRAM OF BIT ERROR OCCURRENCES IN CLIP 1 WITH THE BENCHMARK GAUSSIAN FILTER ATTACK

Thus, we classify the blocks with 3 bit errors according to the position of the errors into the four cases as listed below:

- Case 1: 3 bit errors in the information bits, no bit errors in ECC
- Case 2: 2 bit errors in the information, 1 bit error in ECC
- Case 3: 1 bit error in the information, 2 bit errors in ECC
- Case 4: no bit error in the information, 3 bit errors in ECC

We count the number of blocks that fall into each case with the three clips used in the test. This is shown in Table 1.

TABLE 1 CLASSIFICATION OF BLOCKS BY ERROR POSITIONS

Clip No.	Number of Blocks			
	Case 1	Case 2	Case 3	Case 4
1	67	15	1	1
2	24	12	11	9
3	23	9	1	2
Total	114	36	13	12

As we have said, cases 2, 3 and 4 can further be corrected with the ECC bits from the audio stream. Thus, a total of 61 (36+13+12) blocks out of 175 (35.0%) can be corrected with the ECC bits. This is a significant improvement in the robustness compared to when the ECC is embedded in the video stream only.

Estimation of Robustness Improvement through Calculations

Assuming that the errors are uniformly distributed, we can calculate the ratio of blocks that fall into the four cases listed in the previous section.

For case 1, the number of occurrences where all three error bits occur in the information bits is ${}_{51}C_3$. On the other hand, the number of all occurrences of 3 error bits is ${}_{51}C_3 + {}_{51}C_2 \times {}_{12}C_1 + {}_{51}C_1 \times {}_{12}C_2 + {}_{12}C_3$. Thus the probability that all three error bits occurring in the information bits out of all occurrences of the three bit errors is

$$\frac{{}_{51}C_3}{{}_{51}C_3 + {}_{51}C_2 \times {}_{12}C_1 + {}_{51}C_1 \times {}_{12}C_2 + {}_{12}C_3} = 0.5244$$

Likewise, for case 2, the probability of errors occurring in 2 bits in the information bits and 1 bit in the ECC bits is

$$\frac{{}_{51}C_2 \times {}_{12}C_1}{{}_{51}C_3 + {}_{51}C_2 \times {}_{12}C_1 + {}_{51}C_1 \times {}_{12}C_2 + {}_{12}C_3} = 0.3853$$

For case 3, where 1 bit in the information bits is in error and 2 bits in the ECC bits, the probability of errors is

$$\frac{{}_{51}C_1 \times {}_{12}C_2}{{}_{51}C_3 + {}_{51}C_2 \times {}_{12}C_1 + {}_{51}C_1 \times {}_{12}C_2 + {}_{12}C_3} = 0.0847$$

Finally, for case 4, where all 3 bit errors are in the ECC bits, the probability is

$$\frac{{}_{12}C_3}{{}_{51}C_3 + {}_{51}C_2 \times {}_{12}C_1 + {}_{51}C_1 \times {}_{12}C_2 + {}_{12}C_3} = 0.0055$$

Since only cases 2 to 4 can be corrected with the ECC bits in the audio stream, we can calculate that a total of 0.4755 (47.6%) of the blocks with 3 error bits can be corrected, assuming the bit errors are uniformly distributed. A comparison of the ratio of 3 bit error occurrences that can be corrected is tabulated in Table 2. Simulation results for the three video clips, and for a long video clip with all 3 clips spliced together, is also included.

TABLE 2 COMPARISON OF ROBUSTNESS IMPROVEMENT

Mode	Clip No.	Percent of correctable frames [%]
Simulation	1	20.23
	2	57.14
	3	34.29
	Combined	34.86
Calculated		47.55

As we see, the simulation results in most cases are somewhat lower than the calculations above. Apparently, bit errors are not uniformly distributed. This probably is due to the fact that the watermark is added to the diagonal DCT coefficients, while the Gaussian filters attenuate high horizontal and vertical frequency regions in a concentric circle, and affect

some of the embedded DCT coefficients more than others do. Still, the simple calculation seems to give a relatively good upper-bound estimation of the ratio of frames that can be corrected with the redundant ECC in audio.

Conclusions

We investigate the possibility of watermark robustness improvement by using a parallel watermarking error correction method for audio and video signals, for multimedia streams with short interval editing and/or switching capability. The watermark information with error correction coding (ECC) is split and embedded into audio and video simultaneously. We estimate the possibility of blocks which can be corrected when the watermark is embedded in the video stream with ECC, and the ECC is also redundantly embedded in the audio stream. The redundant ECC in the audio can be used when the ECC in the video includes errors. According to the simple Gaussian filter attack simulation for the case when the Bose-Chaudhuri-Hocquenghem (BCH) code is placed in both video and audio streams, we estimate that for blocks with 3 bit errors (which refers to an overall bit error rate of 5%), approximately 35% of the blocks can further be corrected with BCH (63, 51) compared to when all information is embedded into only the video stream. Thus, the distribution of ECC bits in multiple streams seems to be quite effective.

We would like to test the parallel watermarking on a more state-of-the-art video and audio watermark, and also devise a frame structure where the video frames and audio streams can be easily synchronized so that watermarks can be embedded and detected at the same time frame. We would also like to investigate the robustness enhancement possible using more powerful error correction codes, such as the Reed-Solomon (RS) codes (Reed, 1960) and Low-Density Parity-Check (LDPC) codes (Gallager, 1962). Although these codes can potentially correct much more bits in error, they generally require longer blocks and more redundancy, and would obviously require optimization of the block lengths and allocation strategy in video and audio.

ACKNOWLEDGMENT

The authors thank the members of the lab at the Graduate School of Science and Engineering at Yamagata University for their daily discussions and support.

REFERENCES

- Boney, L., A.H. Tewfik, and K.N. Hamdy. "Digital Watermarks for Audio Signals." Paper presented at the IEEE International Conference on Multimedia Computing and Systems (IEEE Multimedia '96), Hiroshima, Japan, June 17-23, 1996.
- Consumer Digital Video Library. Accessed January 9, 2014. <http://www.cdv1.org>.
- Gallager, R. G. "Low density parity check codes." IRE Trans. Info. Theory, IT-8 (1962): 21-28.
- He, Xing, and Michael S. Scordilis. "Spread Spectrum for Digital Audio Watermarking." In Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks, edited by Nedeljko Cvejić and Tapio Seppänen, 11-49. Hershey, NY: Information Science Reference, 2008.
- Kutter, Martin, and Fabien A. P. Petitcolas. "A fair benchmark for image watermarking systems." Paper presented at the SPIE Electronic Imaging, San Jose, CA, Jan. 25-29, 1999.
- Li, Xin, and Hong Heather Yu. "Transparent and Robust Audio Data Hiding in Cepstrum Domain." Paper presented at the IEEE International Conference on Multimedia and Expo, New York, NY, July 30 - August 2, 2000.
- Morelos-Zaragoza, Robert H. The Art of Error Correcting Coding, Second Edition. Hoboken, NJ: John Wiley & Sons, 2006.
- NHK, Mitsubishi Electric Corp. "A Digital Watermark for Hi-Vision Systems that can Maintain High Quality" July 19, 2011. Accessed January 9, 2014. In Japanese. <http://www.nhk.or.jp/pr/marukaji/m-giju311.html>.
- Petitcolas, Fabien A. P., and Ross J. Anderson. "Evaluation of copyright marking systems." Paper presented at the IEEE International Conference on Multimedia Computing and Systems, Florence, Italy, June 7-11, 1999.
- Piva, A., M. Barni, F. Bartolini, and V. Cappellini, "DCT-based Watermark Recovering without Resorting to the Uncorrupted Original Image." Paper presented at the IEEE Conference on Image Processing, Santa Barbara, CA, Oct. 26 - 29, 1997.
- Reed, I. S., and G. Solomon, "Polynomial Codes over Certain Finite Fields," J. Soc. for Industrial and Applied Mathematics, 8 (1960): 300-304.
- Tanaka, Keisuke, Ogihara Akio, Motoi Iwata, and Akira Shiozaki. "A Correlation-Based Watermark for Movie with Sound Data." Paper presented at the IEICE National Convention, Toyonaka, Osaka, March 21 - 24, 2005. In Japanese.
- Yamaguchi, Yoshimi, Toshio Koga, and Kazuhiro Kondo. "A Watermarking Method for Embedding Short Information Split into Audio and Video Signals." Paper presented at the IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Piraeus-Athens, Greece, July 18-20, 2012.